

# 2020 Census Census Scientific Advisory Committee Post-Enumeration Data Processing Status

**Michael Thieme**

**Barbara LoPresti**

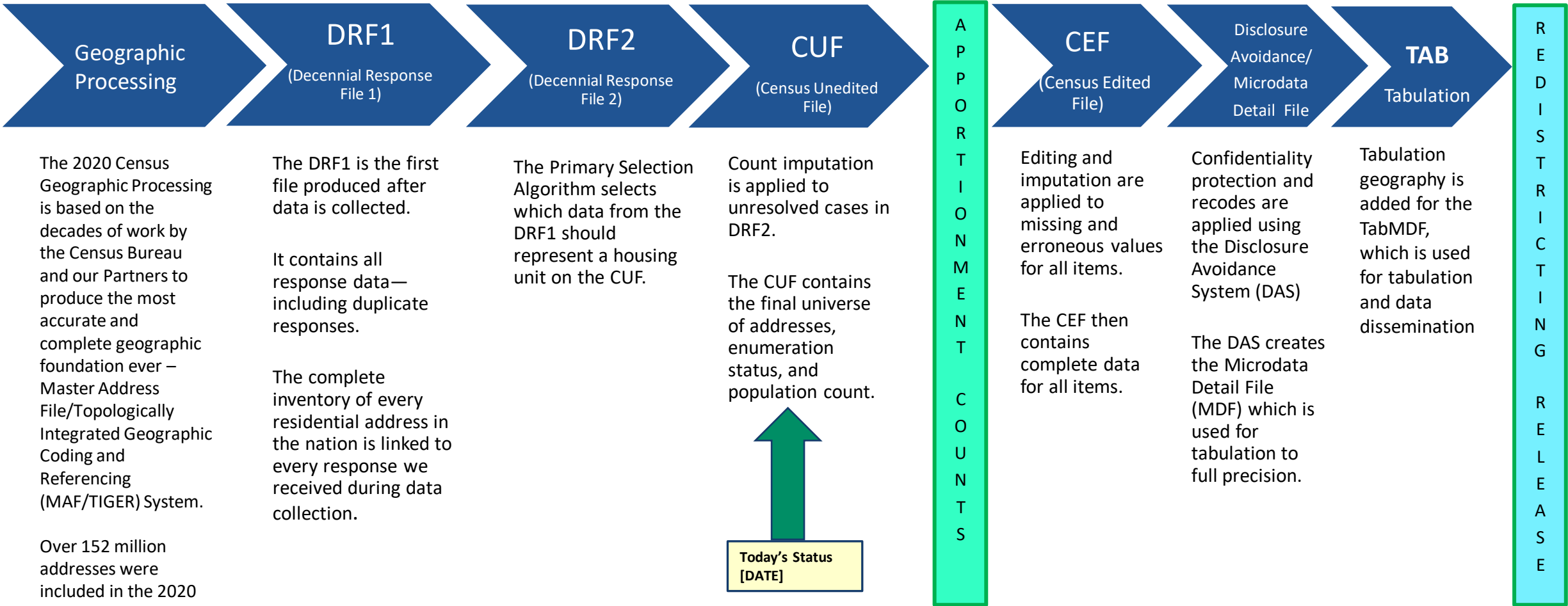
Unclassified/ For Official Use Only

Shape  
your future  
START HERE >

United States<sup>®</sup>  
**Census**  
**2020**

# 2020 Census

## Post Data Collection Processing Flow



# 2020 Census

## Apportionment and Redistricting

### Key Milestones

Processing Step	Start	Finish
Decennial Response File 1 (DRF1)	10/29/2020	12/26/2020
Decennial Response File 2 (DRF2)	12/26/2020	2/24/2021
Census Unedited File (CUF)	2/25/2021	3/26/2021
Apportionment Preparation and Release	3/27/2021	4/30/2021
Census Edited File (CEF)	4/20/2021	6/24/2021
Disclosure Avoidance Application/Microdata Detail File	6/25/2021	7/18/2021
Tabulation File Processing	7/19/2021	8/16/2021
Redistricting Preparation and Release	8/17/2021	9/30/2021

# 2020 Census

## Processing Anomalies

- Webster describes an anomaly as “something different, abnormal, peculiar, or not easily classified”
- Importantly, neither Webster nor the Census Bureau define an anomaly as an error
- Every census has processing anomalies – 2020 is no different
- Anomalies found in processing are not errors in the census, but they can turn into errors if we don’t review and resolve them
- Far from raising concerns, the fact that the Census Bureau's planned process expects and identifies anomalies demonstrates that it is *working*
- Identifying and resolving anomalies is not only expected, it is comforting evidence that quality assurance processes continue to validate our commitment to accuracy

# 2020 Census

## Processing Anomalies

**Total number of anomalies encountered so far in 2020: 33**

### Anomaly Categories

#### **1. Standard problems that arise in processing any large survey (27 of the 33 anomalies encountered)**

- Routine coding anomalies
- Basic errors in processing code
- Processing errors in data handoffs between systems
- Errors in business rules

#### Examples:

- Processing software was miscalculating the age of respondents if those respondents did not include the month and day of their birthday in their response. A simple code correction fixed this, but if it had not been fixed, it may have reduced our ability to match and remove duplicate responses (for example when we get responses from two people in the same household).
- Some responses collected via internet self-response were duplicated in responses from Group Quarters (GQ) addresses. The processing software specified invalidating the Housing Unit responses in favor of the GQ response but did not execute that invalidation properly. A code correction fixed this, but if it had not been fixed the result could have produced an overcount of persons. This is also a good example of addressing responses from multiple modes.

# 2020 Census

## Processing Anomalies

### Anomaly Categories (cont.)

#### 2. **Anomalies resulting from unanticipated respondent action - especially with the impact of the pandemic (5 of the 33 anomalies encountered)**

Example:

- In a small number of colleges and universities the total student population in all dorms was submitted by those institutions for *each* dorm, potentially inflating the population count on those campuses. A code fix enabled distribution of the correct population among the college dorms where this occurred..

# 2020 Census

## Processing Anomalies

### Anomaly Categories (cont.)

#### 3. Anomalies resulting from unanticipated enumerator action (1 of the 33 anomalies encountered)

##### Example:

- During GQ enumeration, there were isolated instances in which some enumerators mistakenly set a GQ rework indicator for one or more persons in the GQ. This invalidated all responses at the GQ. A GQ rework indicator is intended to be set *only* for the entire GQ, not persons within the GQ. A software fix enabled and validated all responses at the GQ.

### ***Key Message***

*Our data processing (including the handling of anomalies) has not shown any critical errors caused by data collection mistakes or omissions that are impossible to fix. We are happy to say we have fixed or are fixing every anomaly that our systems and processes have identified thus far. All processing decisions favored accuracy, not expediency in meeting the schedule.*

# Questions

Shape  
your future  
START HERE >

United States®  
**Census**  
**2020**